Recker, Jan C. and Rosemann, Michael (2008) *Measuring perceived representational deficiencies in conceptual modeling : instrument development and test.* In: 29th International Conference on Information Systems, 12-14 December 2008, Paris, France.

ICIS Proceedings

ICIS 2008 Proceedings

Association for Information Systems *Year 2008*

Measuring Perceived Representaional Deficiencies in Conceptual Modeling: Instrument Development and Test

Jan Recker* Michael Rosemann†

*Queensland University of Technology, j.recker@qut.edu.au
†Queensland University of Technology, m.rosemann@qut.edu.au

MEASURING PERCEIVED REPRESENTATIONAL DEFICIENCIES IN CONCEPTUAL MODELING: INSTRUMENT DEVELOPMENT AND TEST

Completed Research Paper

Jan Recker
Queensland University of Technology
126 Margaret Street, Brisbane QLD 4000, Australia
j.recker@qut.edu.au

Michael Rosemann
Queensland University of Technology
126 Margaret Street, Brisbane QLD 4000, Australia
m.rosemann@qut.edu.au

Abstract

Over recent years, many scholars have studied the conceptual modeling of information systems based on a model of representation formulated by Wand and Weber. This model offers a set of premises about representational deficiencies and their implications for development and use of conceptual modeling in IS practice. Yet, only few examples of empirical research exist. One of the related problems is the lack of validated measurement instruments that could be used in such studies. This paper reports on the development of a valid and reliable instrument for measuring the perceptions that individuals have of the representational deficiencies of conceptual modeling artifacts. We describe a multi-stage approach for instrument development that incorporates feedback from expert and user perspectives. We also report on a field test of the instrument with modeling practitioners. The resulting instrument can be used in future studies that investigate the use of conceptual modeling in IS practice.

Keywords: Representation theory, Bunge-Wand-Weber model, instrument development, perception measurement

INSUFFISANCES DE REPRESENTATION PERÇUES DANS LA MODELISATION CONCEPTUELLE: DEVELOPPEMENT D'UN INSTRUMENT DE MESURE ET TEST

Résumé

Cette étude présente un instrument valide et fiable permettant de mesurer le degré d'insuffisance perçue en termes de représentation des modèles conceptuels. Nous décrivons une procédure multi-étapes de développement d'instruments qui tient compte à la fois des réactions des experts et des points de vue des utilisateurs. L'instrument de mesure est testé sur un échantillon de professionnels de la modélisation.

Introduction

A major task undertaken by systems analysts and designers, workflow engineers, process analysts, and the like, is to develop a model of a real world business domain. The so-called practice of *conceptual modeling* has been found to be a very conducive way of articulating knowledge and perceptions about features of a real world domain that are, or should be, supported by IS applications (Maes and Poels 2007). In line with the popularity of conceptual modeling in practice, conceptual modeling has emerged as a popular and relevant area of research in Information Systems (Wand and Weber 2002). However, by far most of the published research in this space is theoretical, conceptual and/or analytical, with the share of empirical papers in this space reportedly being less than 20 per cent (Moody 2005).

One reason for the lack of empirical research in this area can be seen in the lack of mature theoretical foundations for conceptual modeling (Weber 2003; Siau and Rossi 2008), upon which empirical research strategies could be based and from which testable propositions could be derived. Several approaches have been proposed over time to present theoretical guidance for the development, evaluation and use of conceptual modeling (e.g., Agerfalk and Eriksson 2004; Rockwell and Bajaj 2004). Widespread are the approaches based on theories of ontology (e.g., Milton and Kazmierczak 2004; Guarino and Guizzardi 2006). A prominent example is the work of Wand and Weber (1990; 1993; 1995) towards a theory of representation derived from an ontology defined by Bunge (1977). Wand and Weber's theory, mostly referred to as the Bunge-Wand-Weber (BWW) representation model, has over the last two decades achieved significant levels of scholarly attention and has been used in well over one hundred studies drawing on this theoretical model in research contexts such as modeling language foundations (Wand et al. 1995), model quality measurement (Gemino and Wand 2005), modeling practice (Burton-Jones and Meso 2006; Shanks et al. 2008) or modeling method engineering (Wand 1996). Research based on the BWW representation model has arguably advanced the knowledge and understanding of conceptual modeling and related phenomena; both on a conceptual and empirical level, as even critics concede (Lyytinen 2006). The proven usefulness and ample amount of evidence surrounding this theoretical model in studies on conceptual modeling and related phenomena justifies our interest in this theory, and hence, the research presented in this paper.

While its levels of dissemination are impressive, representation theory is not without its criticisms. More precisely, academics have criticized a lack of empirical validation of the principles that stem from the use of representation theory (e.g., Wyssusek 2006). While there have been some empirical tests on the basis of the BWW representation model (e.g., Bodart et al. 2001; Gemino and Wand 2005), we concur with the critics that IS research has yet to demonstrate – on a large scale – that the premises of representation theory in fact inform conceptual modeling practice, and, moreover, leverage good modeling practice.

The *aim of this paper* is to report on, and discuss, the development of a measurement instrument designed to capture the perceptions that an individual may have towards the main premises of representation theory. The stage of creating valid and reliable measurement instruments for underlying theoretical concepts is vital to any empirical research (Froehle and Roth 2004; Straub et al. 2004). This stage is of even greater importance in research contexts in which new theories have to be adapted for empirical research (Segars and Grover 1993). And indeed, in IS research a large number of measurement instruments designed to tap into the explanatory concepts of underlying theories lack reliability and validity (Straub 1989; Moore and Benbasat 1991; Boudreau et al. 2001; Lewis et al. 2005), thereby further underlining the importance of the measurement instrument development phase.

We proceed in this paper as follows. The next section introduces the main concepts and premises of representation theory and reviews the main works based on this theory. We then report in detail on the procedure we developed and employed for measurement instrument development. Then, results from a confirmatory field study with conceptual modeling practitioners are reported. The paper concludes with a summary of contributions and an outlook to further research.

Background and Theory

Practitioners often use graphical models of the business domain they are concerned with in order to be able to understand and communicate aspects of a complex domain in a simple manner. In IS practice, analysts use so-called conceptual models to represent features of a business domain intended to be supported by an existing or newly built information system. These conceptual models (*scripts*) are specified using a *grammar* (i.e., a set of constructs and rules to combine those constructs), a *method* (i.e., procedures by which the grammar can be used), and apply within an organizational *context* (i.e., the setting in which the modeling occurs) (Wand and Weber 2002).

Conceptual modeling is an essential cornerstone of many IS analysis and design methodologies. This is because conceptual modeling helps articulating knowledge about relevant business domain features and identifying errors at early stages of IS development (Moody and Shanks 2003). Clearly, conceptual modeling is an important practice in IS, however, it is one that is yet to be fully understood. Empirical studies into conceptual modeling practice (Batra and Marakas 1995; Patel et al. 1998; Parsons and Cole 2005; Raduescu et al. 2006; Rosemann 2006) still report on faulty models, lackluster modeling performance, lacking utilization and other mishaps. One reason for this situation is suggested to stem from the lack of widely acknowledge quality frameworks for conceptual modeling(Moody 2005).

Searching for a theoretical basis to aid IS researchers in establishing insights into the nature and characteristics of conceptual modeling, we turn to a theory with a focus on the representational capacities and issues of artifacts nominally ascribed to conceptual modeling. Wand and Weber's (1990; 1993; 1995) representation theory builds upon the philosophical discipline of ontology for presenting a foundational framework for research on conceptual modeling. Generally, ontology studies the nature of the world and attempts to organize and describe what exists in reality, in terms of the properties of, the structure of, and the interactions between real-world things (Bunge 1977). With information systems essentially being human-created representations of real-world systems, Wand and Weber (1990; 1993; 1995) suggest that ontology may help in devising conceptual structures on which modelers can base their representations of these systems. They adopted Bunge's ontology into a model of representation (the BWW representation model), which specifies a set of essential ontological constructs that IS conceptual modeling grammars and methods need to provide representations for. Based on their representation model of ontological constructs, Wand and Weber (1993) developed a set of premises for informing the characteristics of good conceptual modeling practice and good conceptual modeling artifacts (see Figure 1):

- a conceptual model of an information system should be complete, viz., it should not exhibit a deficit of representational constructs that are needed to articulate in a conceptual model all relevant facets of real-world phenomena that a user seeks to describe. Wand and Weber (1993) have labeled this principle *ontological completeness*, and
- a conceptual model of an information system should also be clear, viz., it should provide representations for all relevant facets of phenomena in such a way that the meaning of these representations can unambiguously be interpreted. Wand and Weber (1993) have labeled this principle *ontological clarity* and differentiate three dimensions of clarity, viz., redundancy, overload, and excess. Construct redundancy concerns graphical constructs in a grammar that essentially share the same representational capacity to depict certain aspects of real-world phenomena. Such cases are undesirable as they lead to confusion over which real-world concept can best be represented by a particular graphical construct in a modeling grammar. Construct overload concerns graphical constructs in a grammar that have multiple possible real-world interpretations. Such cases require the user to bring to bear extra-model knowledge in order to understand the capacity in which a given construct is used in a particular modeling scenario. Construct excess concerns graphical constructs in a grammar that, as per theory, do not carry any real-world meaning at all. In such cases it is expected that users will get confused as to their nature and purpose when using these constructs and, hence, they will need mechanisms for further clarification.

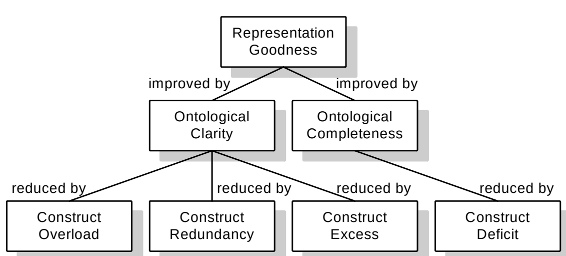

Figure 1. Premises and Fundamental Concepts of Representation Theory (Weber and Zhang 1996)

A wide range of scholarly work has been carried out on the basis of this set of premises to investigate the quality of conceptual modeling (e.g., Rohde 1995; Weber and Zhang 1996; Green and Rosemann 2000; Opdahl and Henderson-Sellers 2002). This work has in common that it shows how conformance to the principles of ontological completeness and ontological clarity affects the quality of conceptual modeling procedures and outcomes. Green (1997), for instance, showed that the occurrence of construct deficit in a conceptual modeling grammar motivated users of the grammar to employ additional grammars in conjunction with the deficient grammar in order to be able to articulate those phenomena that they felt could not be expressed. This in turn resulted in additional cost and complexity of the modeling process. Recker et al. (2006) showed that users tend to avoid excess constructs in order not to decrease the ease of interpreting the resulting model. In another case, Gemino and Wand (2005) showed how the non-conformance to certain principles of ontological clarity negatively affected the levels of domain understanding generated by the conceptual model. Similarly, Bowen et al. (2006) showed how ontologically clearer data models resulted in better query formulation.

Recker (2008) reports on a summarized review of the main works concerning representation theory and its use in various conceptual modeling domains such as traditional, structured, data-oriented, object-oriented, or process–oriented modeling, enterprise systems interoperability, use case specification and reference modeling. However, of the seventy-four works reviewed in (Recker 2008), only 12.2 percent involved empirical studies of some kind. In other words, research on basis of representation theory has created a large body of theoretical knowledge on the characteristics and deficiencies of conceptual modeling. However, we still know only little about how the theory premises inform actual conceptual modeling practice. Moreover, most of the empirical research conducted (e.g., Bodart et al. 2001; Green and Rosemann 2001; Parsons and Cole 2004; Gemino and Wand 2005; Bowen et al. 2006; Recker et al. 2006; Shanks et al. 2008) has studied *objective* measures of conceptual modeling quality, such as levels of understanding generated, model query results, time taken for model creation or interpretation, or number of model errors. IS research has yet to examine another fundamental aspect of conceptual modeling practice, that of user *perceptions* towards the nature, and use, of conceptual modeling artifacts.

The lack of empirical studies may be attributed to a missing instrumental operationalisation of the theory constructs (Weber and Zhang 1996, p. 169). Without the existence of a theoretically sound, rigorously developed and thoroughly tested measurement instrument, research efforts are endangered to lead to mixed and inconclusive outcomes (Lewis et al. 2005). Accordingly, the next section reports on our efforts to derive a measurement instrument for the four main representation theory premises.

Research Approach

In general terms, development of measurement instruments should be carried out in multiple stages and should incorporate validation attempts during creation stages rather than application stages (Nunnally and Bernstein 1994). In fact, instrument validity is considered both a prior and primary validation for all IS empirical research (Straub et al. 2004). Accordingly, we devised an measurement instrument development procedure that incorporates five stages of testing, based on, and extending, the methodological procedures first described by Davis (1986; 1989) and later revised by Moore and Benbasat (1991). We selected these methodological procedures over others (e.g., de Vaus 2001; Lewis et al. 2005), because they explicitly pay attention to the objective that the resulting instrument, while being developed for a particular research purpose, is general enough to allow for a wider uptake in other related studies (Moore and Benbasat 1991, p. 194). We deemed this to be of particular pertinence to our research, as previous research has shown that Wand and Weber's representation theory facilitates insights into a large variety of phenomena ascribed to conceptual modeling (see previous section). Figure 2 depicts the overall procedure and gives information about the research method deployed in each stage as well as references to the relevant literature.

As shown in Figure 2, the first stage of the development procedure is *item creation*, which is concerned with specifying the theoretical constructs for which measurement items are to be developed, and to derive pools of candidate items for each construct. The next stage is *substrata identification*, the purpose of which is to sort the candidate items into meaningful separate domain sub-categories. This step is important for identifying relevant dimensions of content validity as well as to display to convergent and discriminant validity. The third stage is *item identification*, the purpose of which is to identify from the pool of candidate items a revised set of items that show good potential for high content validity. The fourth stage is *item selection and revision*, the purpose of which is to re-specify and further condense the set of candidate items as well as to get an initial indication of reliability and validity. Overall, the four stages are primarily concerned with establishing content validity of the measurement instrument, following the observation of Straub et al. (2004) that content validity, while hard to assess, is of key importance to developing new, good measures, especially in the absence of prior empirical practice. The last step of our procedure then is *instrument validation*, which is concerned with obtaining statistical evidence for validity and reliability of the developed measurement items, i.e., to assess quantitatively all other mandatory criteria for instrument validation. The following subsections describe each of the steps in more detail.

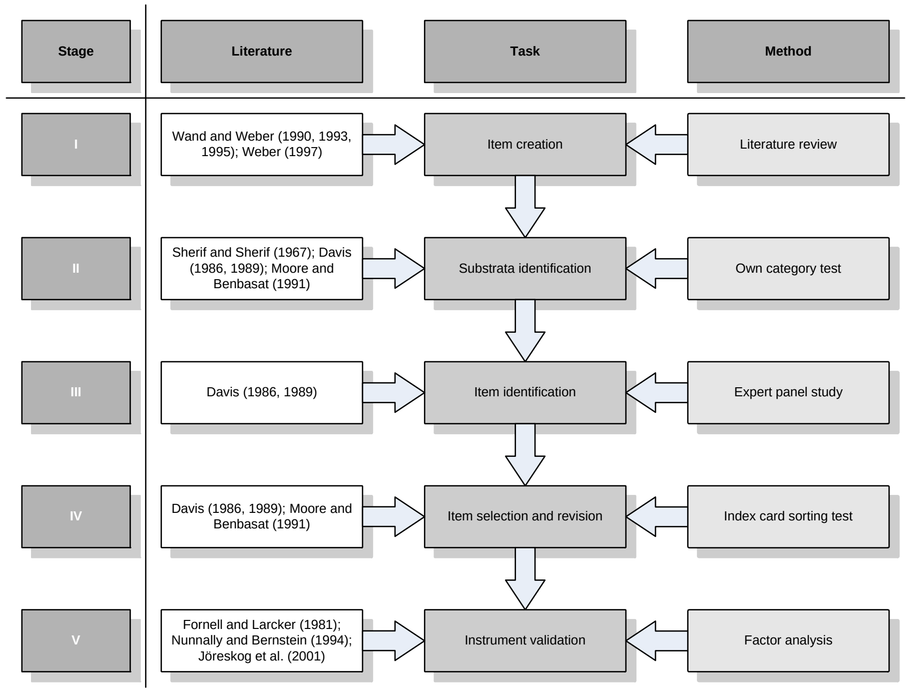

Figure 2. Instrument Development Process

In the following, we report on how we carried out the measurement instrument development process. In doing so, we refer to one specific artifact associated with conceptual modeling, that of a conceptual modeling grammar. In particular, our discussion uses the example of the BPMN modeling grammar (BPMI.org and OMG 2006) for the graphical modeling of process-aware information systems. While this limits the scope of our research effort, we have no reason to believe that our findings cannot be generalized and adopted in studies of other conceptual modeling artifacts (such as other grammars, methods or scripts).

Instrument Development Process

Stage 1: Item Creation

A sound specification of the theoretical constructs to be measured is the origin of any operationalisation (Stone 1981). In our case, representation theory offers four constructs that form the basis of the principles of ontological completeness and clarity, these being construct deficit, construct redundancy, construct overload and construct excess. Typically, the extent to which modeling artifacts (such as modeling grammars) exhibit a deficiency in any of these four situations can be established by means of representational analysis (e.g., Wand and Weber 1993; Rosemann et al. forthcoming). However, it needs to be considered that the four measures of representation theory do not necessarily affect actual conceptual modeling practice. Representational deficiencies (such as high levels of construct deficit, redundancy, overload or excess) do not in all cases imply a practically observable disadvantage or issue (Gemino and Wand 2005). In other words, representational deficiencies are not necessarily perceived as such. To be able to account for a potential perception 'gap', we consider the *secondary* attributes of modeling artifacts (i.e., user perceptions towards ontological completeness and clarity), instead of the *primary* attributes of modeling artifacts (i.e., their analytically established levels of ontological completeness and clarity).

Our decision follows the arguments provided by Downs Jr. and Mohr (1976) who concede that secondary qualities of an artifact, i.e., an individual's perceptions of its primary qualities, determine the formation of an attitude towards it and thereby associated usage behavior. In fact, it is most often not the actual qualities of an artifact that will influence an attitude towards it, but rather the perception of the qualities by the user (Moore and Benbasat 1991). Consider the case of a product that is annotated with a certain price. A purchase decision will not be made on the basis of the actual (i.e., primary) price attribute but rather on whether an individual perceives the price to be reasonable or too expensive (i.e., the secondary attribute). In the context of conceptual modeling, several studies found that a representational (dis-) advantage does not in all cases imply a practice or observable (dis-) advantage (e.g., Gemino and Wand 2005; Burton-Jones and Meso 2006; Recker et al. 2006). This means that it must be exploited whether users of conceptual modeling artifacts perceive theorized primary qualities and/or deficiencies as such. There are two aspects to consider:

- whether or not a modeling artifact has a perceived deficiency. For instance, whether users of a conceptual modeling grammar experience a deficit of construct and/or a number of constructs that representation theory informs us are unclear, and
- whether or not these artifact deficiencies have a practical or observable impact on users using the artifact. For instance, whether users of a modeling grammar perceive that they are unable to articulate a certain phenomenon because there is construct deficit in the grammar.

Taking these considerations into account, we adopted the original construct definitions (Wand and Weber 1990; 1993; 1995; Weber 1997) as follows:[1]

- Perceived construct deficit (CD): The extent to which a conceptual modeling grammar user perceives the grammar to have a deficit of constructs that (s)he would require to describe all real-world phenomena that (s)he seeks to have represented in a conceptual model.
- Perceived construct redundancy (CR): The extent to which a conceptual modeling grammar user perceives the grammar to provide more constructs than required to describe a single real-world phenomena that (s)he seeks to have represented in a conceptual model.
- Perceived construct overload (CO): The extent to which a conceptual modeling grammar user perceives the grammar to provide constructs that can each be used to describe more than one single real-world phenomena in a conceptual model.
- Perceived construct excess (CE): The extent to which a conceptual modeling grammar user perceives the grammar to provide constructs that do not describe any relevant real-world phenomena in a conceptual model.

Forthcoming from the specification of the construct definitions is the need to pursue a set of appropriate measurement items for these constructs. The objective of generating candidate item pool is to ensure content validity (Moore and Benbasat 1991). To that end, the representation theory-related literature was closely examined to generate pools of candidate items for each of the four constructs. In doing so, we referred to the use of the Spearman-Brown Prophecy formula in Davis's (1989) study of technology acceptance as an indication of how many items to create. He suggests that at least ten items per construct are needed to achieve reliability levels of at least .8. As per specification of the candidate items, Ajzen and Fishbein's (1980) suggestions were followed to include in the definition of the items the actual behavior (e.g., using a conceptual modeling grammar to create models), the target at which the behavior is directed (e.g., ERMs as a specific conceptual modeling grammar), the context in which the behavior occurs (e.g., for database specification tasks) and, where possible, a time frame (e.g., the most recent data specification initiative). Accordingly, we included in our item specification the example of a palpable conceptual modeling grammar, in our case the BPMN grammar for process modeling (BPMI.org and OMG 2006). This was done, to make the item more tangible and understandable as well as to refer to a specific context of behavior, viz., conceptual modeling for specifying process-aware information systems. It should be noted, however, that we use the example of BPMN for illustration purposes only. In pursuing our objective of generating measurement instruments we ensured that our resulting items are applicable to any conceptual modeling domain and/or artifact. For instance, during all stages of the measurement item development procedure outlined below, we worded the candidate items as general as possible. This approach, in turn, allows felloe researchers to use these items as a 'template' that can be adopted to a wide range of modeling artifacts (e.g., methods, tools, grammars, scripts), modeling domains (e.g., object, process, data) or even different types of representational deficiencies.

[1] Note that during the expert panel test reported later in this paper, the participating IS scholars were asked for feedback on the construct definitions. The construct definitions shown incorporate the suggestions voiced.

Due to word limitations we omit an in-depth discussion of the development procedure for all four theory constructs (CD, CR, CO, CE) and instead report on illustrative examples (taking the case of perceived construct deficit). Table 1 gives the initial item pool for perceived construct deficit. Item creation for the remaining three constructs was accomplished in a similar fashion and results can be obtained from the contact author upon request.

Table 1. Initial Candidate Items for Perceived Construct Deficit

Item No	Item Definition
CD1	BPMN lacks capacities for representing certain real-world phenomena in process models.
CD2	BPMN’s support for the representation of certain real-world phenomena in process models is deficient.
CD3	BPMN users lack capacities to represent certain real-world phenomena in process models.
CD4	BPMN does not provide a sufficient number of constructs for representing certain real-world phenomena in process models.
CD5	BPMN users cannot sufficiently represent certain real-world phenomena in process models using the constructs provided.
CD6	The constructs needed to represent certain real-world phenomena in process models are not fully provided by BPMN.
CD7	BPMN does not contain all constructs needed to represent certain real-world phenomena in process models.
CD8	The extent to which BPMN provides constructs that are needed for representing certain real-world phenomena in process models is not sufficient.
CD9	There are not enough constructs in BPMN to represent certain real-world phenomena in process models.
CD10	Certain real-world phenomena cannot be represented in process models using BPMN.

Stage 2: Substrata Identification

Forthcoming from the generation of an initial pool of candidate items is the establishment of construct validity of the candidate items, in particular the display of convergent and discriminant validity. To that end, we a procedure called ‘own category test’ (Sherif and Sherif 1967). In this test, a panel of judges is asked to sort candidate items into a number of construct categories so that the statements within a category are most similar in meaning to each other and most dissimilar in meaning from those in other categories. The categories are also to be labeled. The labels are then used to assess whether the identified substrata appropriately reflect the item’s intent. Categorization provides a simple yet powerful indicant of cluster similarity that helps to reflect on the domain substrata for each construct and thus to assess coverage and representativeness of the items. Incorporating labeling into this procedure further minimizes the risk of interpretational confounding (Burt 1976), which occurs when study participants assign to a measurement item a meaning other than the a priori intended.

In our case, a panel of four recognized academics with a strong track record in studies using the BWW model was asked to sort the initial candidate items into construct categories. Panel members were selected from the overall population of IS researchers that have published articles on studies that build upon representation theory. An overview of these researchers’ work can, for instance, be found in (Recker 2008). Each panel member was contacted individually and received instructions including a sample categorization test. This step was done to ensure the mechanics of the test were fully understood by the participating panel members.

The panel members used at most four categories in which they placed the candidate items. Table 2 gives the results for perceived construct deficit. The received categorization data was cluster analyzed. This was done to identify domain substrata of the primary theoretical constructs. Measurement items should tap into each of the identified substrata in order to provide a comprehensive coverage of the construct’s domain content with reduced chance of measurement error.

In performing the clustering of the categories obtained from the panel members, two coders separately clustered the categories, then met to defend their clusters and created a joint draft, thereby reducing subjectivity in the coding procedure. We calculated inter-coder reliability by means of Cohen's Kappa (Cohen 1960). We obtained a Kappa value of .72 when comparing the individual clustering results, indicating sufficient reliability (Straub et al. 2004). We found two very clear and strong clusters to emerge across all four theoretical constructs, one tapping into 'grammar deficiency' and the other tapping into 'impact of a deficiency on the user of the grammar'. In hindsight, this is not a surprise given that the theoretical premises of the BWW representation model incorporate exactly these two aspects. The identification of the clusters serves for all subsequent as a reference point to ensure adequate coverage of the domain substrata for each of the theoretical constructs.

Table 2. Substrata Identification Results for Perceived Construct Deficit

Item No	Rater1	Rater2	Rater3	Rater4
CD1	-	number/ suitability of constructs	lack of support	Missing constructs
CD2	deficit	number/ suitability of constructs	lack of support	Lacking capacities for users
CD3	deficit	user capacity	lack of capacities provided to user	Lacking capacities for users
CD4	specializations	number of constructs	lack of constructs	Missing constructs
CD5	number of constructs	number/ suitability of constructs	lack of capacities provided to user	Lacking capacities for users
CD6	specializations	number of constructs	lack of constructs	Missing constructs
CD7	completeness	number/ suitability of constructs	lack of constructs	Missing constructs
CD8	deficit	suitability of constructs	lack of constructs	Missing constructs
CD9	deficit	number of constructs	lack of constructs	Missing constructs
CD10	completeness	lack of capability	lack of support	Lacking capacities for users

A second step was to assess whether panel members consistently placed the same candidate items in these clusters. Following the recommendations of Moore and Benbasat (1991) we demonstrated reliability of the cluster scheme by assessing the percentage of items placed in the target cluster across all panel members, which indicates the degree of inter-judge agreement. Also, the items that obtained high placement percentages across the panel show high potential for high construct validity and reliability. The four panel members agreed in at least 75% of their item placements when mapped to the two coded substrate deficiency and impact, which in turn increases confidence in the validity and reliability of the cluster scheme.

Stage 3: Item Identification

The goal of the item identification stage was to establish differences in content validity between the candidate items in order to be able to drop items that show little potential for high validity. To that end, a panel of representation theory experts was asked to assess, on a 7-point scale, the correspondence between the candidate items and the definitions of the constructs they are intended to measure. This step followed the procedures firstly documented by Davis (1989).

The expert panel consisted of thirteen recognized IS academics with varying yet strong track records in studies on basis of the BWW model. By forming a panel of representation theory experts, higher potential for correctly assessing the validity of candidate items was achieved. In addition, we sought informal, qualitative feedback on our development procedure. Overall, twenty-three contributors to representation theory in IS research were identified and contacted, and a response rate of 57% (13) was obtained. Each panel member received instructions, including a sample test as well as construct definitions and candidate items. Additional space was allowed for extended reasoning and feedback.

The responses of the panel members were averaged and ranked to obtain an order of candidate items with respect to their content validity. This was done to identify potential candidates for elimination. In eliminating items, however, it had to be considered whether the remaining item pool contains appropriate representativeness of the domain substrata (deficiency and impact) of the theoretical construct (Bohrnstedt 1970). Hence, in analyzing the results attention was paid to the results of the categorization task (see previous section) in order to identify domain substrata of which the item pool may have excessive, or inadequate, coverage. As an example, CD1 received a relative good ranking but was found not to resemble any of the identified two domain substrata. However, given its high ranking, item CD3 was reworded to incorporate some of the content of CD1. Similarly, CD3 was retained (but reworded) even though it received low rankings, with the rationale being that it was one of only two items (CD3 and CD10) that clustered appropriately in the ‘impact’ substratum.

Overall, this step resulted in an ordered ranking of potential construct validity of the candidate items. This ranking was then used to eliminate items that demonstrated low potential for validity (e.g., items CD2, CD5). Both the ranking and categorization data for the CD construct are summarized in Table 3.

Table 3. Expert Panel Results for Perceived Construct Deficit

Item No	Ranking average	Standard deviation	Rank	Cluster	Placement ratio	**Decision** / New Item wording
CD1	6.077	0.954	1	-	-	**Dropped**
CD2	5.077	1.382	5	-	-	**Dropped**
CD3	3.923	1.847	10	Impact	75 %	**Retained**: BPMN does not provide me with sufficient capacities for representing certain real-world phenomena in process models.
CD4	5.154	1.819	3	Deficiency	75 %	**Retained**: BPMN does not provide types of constructs to represent certain real-world phenomena in process models.
CD5	5.000	1.291	6	-	-	**Dropped**
CD6	5.692	1.601	2	Deficiency	75 %	**Retained**: BPMN could be improved by adding new constructs for representing certain real-world phenomena in process models.
CD7	5.154	2.340	3	Deficiency	100 %	**Dropped**
CD8	4.462	1.854	8	Deficiency	75 %	**Dropped**
CD9	4.923	1.891	7	Deficiency	100 %	**Dropped**
CD10	4.385	1.609	9	Impact	75 %	**Retained**: I often cannot use BPMN to adequately represent certain real-world phenomena in process models.

In order to reach a decision on item identification, the ranking data together with the categorization data as well as feedback from the panel were considered. Hence, as shown in Table 3, some items were reworded based on suggestions (e.g., CD4), others were reworded to incorporate content from items that exhibited high rankings but failed to cluster appropriately (e.g., CD1 content was incorporated into CD3), and some items (e.g., CD10) were reworded with the view to improving content validity ranking whilst maintaining the clustering into one of the two domain substrata. In addition, informal qualitative feedback from the panel members was used to improve the wording of some of the items (e.g., CD6).

Stage 4: Item Selection and Revision

The fourth stage of the process was to revise the reduced set of candidate items to a final set of ‘high potential candidate items’ and to improve their potential validity and reliability. An appropriate procedure for this type of task is the index card sorting test established by Moore and Benbasat (1991) following the initial idea of Davis (1989).

In the sorting test, a panel of judges was randomly given the items printed on index cards and asked to sort these cards into categories. In different rounds of this test, the categories in which the items are to be sorted into were either given to the panel of judges or not. Moore and Benbasat (1991) recommend four rounds of sorting, each with a different panel, and alternating between given and not-given categories. This recommendation was adopted in our study. This means that in rounds one and three, judges independently had to make up categories for the items, which were later compared to the originally intended categories. In rounds two and four, judges were asked to sort items into categories given to them, and to identify items that are ambiguous or indeterminate.

We selected panel members that were not familiar with the underlying theory of representation. This was done to obtain an indication of how well the measurement items would be understood by practitioners unaware of the theoretical considerations. Overall, sixteen judges, including professional staff, consultants, analysts and post-graduate students, participated over the four rounds, none of them familiar with the study domain. By including members with different theoretical and practical expertise we sought to incorporate adequate proxies for varying types of modeling practitioners. In each round, the panel size varied between three and five members. Each panel of judges was gathered together in a face-to-face setting to explain the intent and mechanics of the test. Two trial sorts were conducted prior to the actual sorting to increase familiarity with the procedure.

To assess the reliability of the sorting conducted by the judges, two measurements were established. Table 4 summarizes coding reliability results in terms of *placement ratio summaries* (e.g., Moore and Benbasat 1991) across all four rounds of sorting and also displays inter-judge agreements measured using Cohen's Kappa (Cohen 1960). Round-by-round revisions helped improve reliability so that, in the end, generally recommended Kappa levels of .7 (Straub et al. 2004) were met.

Table 4. Coding Results from Index Card Sorting Test

Measure	Round 1	Round 2	Round 3	Round 4
Average Kappa	*0.34*	*.64*	*.43*	*.80*
Placement ratio summary				
Construct deficit	100.00%	100.00%	87.50%	100.00%
Construct redundancy	58.33%	87.50%	50.00%	66.67%
Construct overload	16.67%	62.50%	66.67%	80.00%
Construct excess	33.33%	75.00%	87.50%	70.00%
Average	*52.08%*	*81.25%*	*72.92%*	*79.17%*

From Table 4 it can be observed how results vary between Rounds 1,3 and 2,4, respectively. This situation was to be expected given that in rounds 1 and 3 judges were not given item categories, which made it harder to categories the items correctly. There is no definite guidance available as to when to stop this exercise. It is advisable, however, to proceed with the rounds of sorting until inter-judge agreement meets acceptable reliability levels for the rounds in which categories were provided as well as for those rounds where categories were not provided. To that end, Landis and Koch (1977) suggest different Kappa thresholds, from fair and substantial to excellent, that could be used as a basis for decision making.

After each round, each set of items was inspected and, if deemed necessary, reworded. Some items (e.g., CD1, CO4) that were repeatedly misplaced (and thus showed only little potential for high validity) were dropped. Appendix 1 gives an overview of the resulting top three candidate items for each construct after these four stages of instrument development. We selected three items per construct for the pragmatic reason of keeping the overall instrument concise and short and for the technical reason of maintaining the minimum number of items required for appropriate measurement model estimation (Jöreskog and Sörbom 2001). At least one measurement item each taps into the 'deficiency' substratum of the construct and the other into the 'impact' substratum. Also note that Appendix 1 displays the items in their final wording after pre- and pilot-tests (see Stage 5 below).

Stage 5: Instrument Validation

Up to this point, the scale development procedure described is more of a qualitative analysis than a rigorous statistical test of validity and reliability of the scales. Of course, without full-scale tests of the complete measurement instrument there is no way of establishing beyond concern whether or not the items in fact measure what they intend to measure. However, we believe that the procedures applied so far have been found very helpful in determining ‘good’ candidates for measurement items.

To prove our ‘hunch’, the next step was to conduct a field test of the measurement instrument developed with a sample of conceptual modelers. The objective was to ensure that the mechanics of compiling the measurement instrument had been adequate and to obtain formal measures for reliability and validity by means of a factor analysis. To that end, we operationalised the candidate items using the example of the BPMN modeling grammar and the representational deficiencies it was found to exhibit (see Appendix 1).

As can be seen from Appendix 1, instead of measuring user perceptions of the representational deficiencies of the grammar as a whole (e.g., through items such as “I believe that the BPMN grammar is redundant”), we decided to adopt a feature-centric view (Jasperson et al. 2005), viz., to operationalize the measurement instrument on a grammar construct-level. This step was done for three main reasons. First, this step allowed us to not to treat a modeling grammar as a black box but instead to examine the actual features of a grammar, viz., the nature and type of its graphical constructs. Second, we are concerned with developing a measurement instrument that can be used to test the premises of representation theory. Representation theory allows researchers to speculate about the nature and implications of grammar constructs and thus we thought it advisable to derive measurement items that operate on the same construct level as the original propositions. Third, adopting this feature-centric view also allowed us to instantiate the generically worded measurement items (e.g., “The BPMN modeling grammar does not provide sufficient symbols to represent certain real-world phenomena in process models”) using clear examples with specific wording (e.g., “The BPMN modeling grammar does not provide sufficient symbols to represent business rules in process models”). In turn, the final items will be more understandable to end users participating in field tests. This being said, we do not foresee much problems in operationalizing, and using, the measurement instrument on a grammar level. This approach would allow scholars to examine the general perceptions of grammar users about deficiencies of the grammar, and to contrast these perceptions to the analytically established, theory-based suggested deficiencies. In fact, it may well be the case that these two sets of deficiencies (perceived versus ‘objective’) do not overlap.

In our case, we refer to the study described in (Recker et al. 2005; 2006) that developed overall nine propositions about the types of representational deficiencies pertaining to various constructs of the BPMN grammar. From this analysis, we used eight representational deficiencies they identified (two per type of deficiency) to operationalise and implement the above measurement items. For instance, one of the proposition of Recker et al. (2005; 2006) speculates construct overload in the Lane construct in BPMN. We thus were able to use this specific example to word the measurement items accordingly for the field test. Appendix 1 lists the all items in the measurement instrument as used in the final field test.

Before administering the field study we ran a pre-test and a pilot test. In the pre-test four academics with knowledge of the study were asked to complete a paper-based version of the survey instrument in face-to-face meetings. During survey completion, notes were taken based on comments received. After instrument revision, the measurement instrument was pilot-tested with a sample of 41 post-graduate students with knowledge of the target grammar. After exploratory factor analysis, changes were made to the measurement instrument and to the items that indicated problems in meeting required validity and reliability thresholds.

Data collection was conducted through survey research, which is a typical method for testing models and validating scales in IS (Grover et al. 1993). The population of interest for this study included conceptual modelers who have knowledge of a certain modeling grammar, viz., BPMN. To that end, a web-based survey instrument was crafted and announced via modeling practitioner forums and online groups. Overall, 590 usable results were obtained over a period of four months during 2007. Participants were selected using a judgmental sampling technique following the guidelines of Sivo et al. (2006). Chi-square tests of key demographic variables showed no significant differences in responses between early and late respondents, indicating the absence of non-response bias.

Reliability and validity for the four measurement instruments (CD, CR, CO, CE) was assessed via confirmatory factor analysis (CFA) techniques implemented in LISREL Version 8.80 and SPSS Version 13.0. Each measurement item was modeled as a reflective indicator of its hypothesized latent construct. All constructs were allowed to co-vary in the CFA model. Table 5 gives the results from the item validation.

Table 5. Item Validation Results

Construct	Item	Item loading	t-statistic (for λ)	Cronbach's α	ρ_c	AVE
CD1 [Business Rules]	CD1_1	0.916	26.957	0.868	0.847	0.916
	CD1_2	0.856	27.075			
	CD1_3	0.889	26.958			
CD2 [Process structure]	CD2_1	0.939	37.226	0.925	0.883	0.938
	CD2_2	0.874	33.499			
	CD2_3	0.889	37.224			
CR1 [Transformations]	CR1_1	0.832	27.435	0.911	0.863	0.928
	CR1_2	0.891	29.785			
	CR1_3	0.906	27.434			
CR2 [Events]	CR2_1	0.922	31.012	0.900	0.877	0.934
	CR2_2	0.939	33.157			
	CR2_3	0.880	31.013			
CO1 [Pool]	CO1_1	0.811	30.349	0.917	0.855	0.926
	CO1_2	0.871	31.148			
	CO1_3	0.842	30.358			
CO2 [Lane]	CO2_1	0.878	31.853	0.914	0.864	0.932
	CO2_2	0.798	33.617			
	CO2_3	0.828	31.938			
CE1 [Off-page connector]	CE1_1	0.927	42.984	0.967	0.924	0.961
	CE1_2	0.952	43.603			
	CE1_3	0.920	43.955			
CE2 [Multiple Instances]	CE2_1	0.943	27.453	0.945	0.907	0.951
	CE2_2	0.917	27.145			
	CE2_3	0.913	28.813			

Based on the data obtained and displayed in Table 5, four tests can be performed. Regarding uni-dimensionality, Cronbach's α should be greater than or equal to .7 to consider items to be uni-dimensional and to be combinable in an index (Nunnally and Bernstein 1994). Table 5 shows that all constructs have α of at least .8, thereby meeting the test of uni-dimensionality.

Reliability refers to the internal consistency of a measurement instrument. Again, the most widely used test for internal consistency is Cronbach's α, which – as a measure of reliability – should be higher than .8 (Nunnally and Bernstein 1994). A second test uses the composite reliability measure ρ_c, which represents the proportion of measure variance attributable to the underlying trait. Scales with ρ_c greater than .5 are considered to be reliable (Jöreskog et al. 2001). Table 5 shows that all constructs obtained α of at least .8 and also meet the required ρ_c cut-off value of .5. These results suggest adequate reliability.

Convergent validity tests if measures that should be related are in fact related. Convergent validity can be tested using three criteria suggested by Fornell and Larcker (1981): (1) all indicator factor loadings (λ) should be significant and exceed .6, (2) construct composite reliabilities ρ_c should exceed .8 and (3) average variance extracted (AVE) by each construct should exceed the variance due to measurement error for that construct (i.e., AVE should exceed 0.50). Table 5 shows that all factor loadings λ are significant at $p < .001$ (see the reported t-values) and exceed the recommended threshold of .6. In terms of composite reliabilities, Table 5 shows that ρ_c exceeded .8 for all constructs. As reported in Table 5, AVE for each construct is higher than .9 suggesting that for all constructs

AVE well exceeded the variance due to measurement error. Overall, it is concluded that the conditions for convergent validity were met.

Discriminant validity tests if measures that should not be related are in fact unrelated. Fornell and Larcker (1981) recommend a test of discriminant validity, where the AVE for each construct should exceed the squared correlation between that and any other construct considered in the factor correlation matrix (not shown due to lack of space). In the present study, the largest squared correlations between any pair of constructs within the measurement model was .783 (between CO1 and CO2), while the smallest obtained AVE value was .916. These results suggest that the test of discriminant validity was met.

Conclusions

The instrument development research outlined in this paper provides several contributions. First, this paper reported on the process of rigorously creating an overall instrument to measure the perceptions of individuals towards representational deficiencies of conceptual modeling artifacts. The procedure described ensures high levels of confidence in developing content validity whilst also establishing construct validity and reliability of the instrument. The procedure employed was found to be helpful and rigorous and should motivate researchers to adopt this design in related empirical studies. Our confirmatory factor analysis shows that indeed the developed measurement instrument demonstrates strong reliability and excellent validity.

Second, the results obtained, i.e., the resulting instrument, can be used in various studies to investigate how individuals perceive deficiencies in the representational capacity of conceptual modeling artifacts. We identify a number of research streams in conceptual modeling that could benefit from the use of our instrument. For example, studies of the post-adoptive usage behavior of conceptual modeling artifacts (e.g., Recker 2007) can study user perceptions of the nature of a conceptual modeling artifact (for instance, a grammar and its constructs) and examine the effects that these perceptions have on usage intentions or usability beliefs. Similarly, the ongoing stream of research that investigates the quality of conceptual models (e.g., Maes and Poels 2007) could use the instrument to understand how user perceptions influence the quality of, and ultimate usage, of a model produced. Also, studies on the critical success factors of conceptual modeling projects (e.g., Bandara et al. 2005) may want to use our instrument to understand how project stakeholders perceive the nature and characteristics of the modeling artifacts employed in these projects, and how these artifacts contribute – or not – to overall project success. Of course, we can also envisage our instrument to be used in empirical studies of other sorts – for instance, of systems alignment processes (Rosemann et al. 2004) or business interoperability (Green et al. 2005).

Third, our paper aims to contribute to a heightened awareness of the need for well-developed measurement instruments, not only in the space of conceptual modeling but for all IS domains. We hope our results are stimulating for other researchers to tackle the difficult task of measurement instrument development. A major intent of this paper accordingly was to suggest procedural guidelines to those engaging in similar endeavors to arrive at well-suited and highly useful tools for empirical research.

No single measurement instrument can meet all needs. The measurement instrument developed in this research addresses one very important aspect of IS practice: the perceptions of users about the characteristics and deficiencies of conceptual modeling artifacts they use for IS analysis and design. While perceptions and beliefs are fundamental to understanding actual behaviors (Downs Jr. and Mohr 1976; Sveiby 1997), clearly, there are some concerns with the approach taken. First, the instrument taps into secondary qualities of conceptual modeling artifacts but not their primary qualities. Analytical studies could be carried out to examine objectively attributable qualities of modeling artifacts and to contrast these with the user perceptions of these qualities. Second, the instrument uses as a conceptual basis findings from representational analyses of conceptual modeling artifacts – which potentially induces subjective bias in the analysis (Rosemann et al. forthcoming). Third, while we took all possible precautions to develop the measurement instrument to be as general as possible, in our development process we used the example of a specific grammar as a conceptual modeling artifact. Yet, while we lack evidence for this claim, we would not expect major difficulties in adopting our procedures or the final instrument to the case of other grammars, methods, scripts or tools. In conclusion, we believe that we have contributed with our final instrument and we hope that it can serve fellow scholars as a useful tool for the study of conceptual modeling practice.

Appendix

Appendix 1. Final Instrument		
Theory Construct	No	Item Definition
CD1 [Business Rules]	CD1_1	BPMN does not provide sufficient symbols to represent business rules in process models.
	CD1_2	BPMN could be made more complete by adding new symbols for representing business rules in process models.
	CD1_3	I often cannot use BPMN to adequately represent business rules in process models.
CD2 [Process structure]	CD2_1	BPMN does not provide sufficient symbols to represent the process structure and decomposition in process models.
	CD2_2	BPMN could be made more complete by adding new symbols for representing the process structure and decomposition in process models.
	CD2_3	I often cannot use BPMN to adequately represent the process structure and decomposition in process models.
CR1 [Transformations]	CR1_1	I often have to choose between a number of BPMN symbols to represent one kind of a transformation in a process model.
	CR1_2	BPMN often provides two or more symbols that can be used to represent the same kind of transformation in a process model.
	CR1_3	In a process model, one kind of a transformation can often be represented by different BPMN symbols.
CR2 [Events]	CR2_1	I often have to choose between a number of BPMN symbols to represent one kind of an event in a process model.
	CR2_2	BPMN often provides two or more symbols that can be used to represent the same kind of event in a process model.
	CR2_3	In a process model, one kind of an event can often be represented by different BPMN symbols.
CO1 [Pool]	CO1_1	I often have to provide additional information to clarify the context in which I want to use the Pool construct in a process model.
	CO1_2	The Pool construct in BPMN can have more than one meaning in a process model.
	CO1_3	I often use the Pool construct to represent more than one real-world phenomena in a process model.
CO2 [Lane]	CO2_1	I often have to provide additional information to clarify the context in which I want to use the Lane construct in a process model.
	CO2_2	The Lane construct in BPMN can have more than one meaning in a process model.
	CO2_3	I often use the Lane construct to represent more than one real-world phenomena in a process model.
CE1 [Off-page connector]	CE1_1	The Off-page Connector construct does not have any real-world meaning in a process model.
	CE1_2	I often cannot precisely articulate the meaning of the Off-page Connector construct in a process model.
	CE1_3	The Off-page Connector construct does not represent any relevant real-world phenomena in a process model.
CE2 [Multiple Instances]	CE2_1	The Multiple Instances construct does not have any real-world meaning in a process model.
	CE2_2	I often cannot precisely articulate the meaning of the Multiple Instances construct in a process model.
	CE2_3	The Multiple Instances construct does not represent any relevant real-world phenomena in a process model.

References

Agerfalk, P.J. and Eriksson, O. "Action-oriented Conceptual Modelling", *European Journal of Information Systems* (13:1) 2004, pp. 80-92.

Ajzen, I. and Fishbein, M. *Understanding Attitudes and Predicting Social Behavior*, Prentice Hall, Englewood Cliffs, New Jersey, 1980.

Bandara, W., Gable, G.G., and Rosemann, M. "Factors and Measures of Business Process Modelling: Model Building Through a Multiple Case Study", *European Journal of Information Systems* (14:4) 2005, pp. 347-360.

Batra, D. and Marakas, G.M. "Conceptual Data Modeling in Theory and Practice", *European Journal of Information Systems* (4) 1995, pp. 185-193.

Bodart, F., Patel, A., Sim, M., and Weber, R. "Should Optional Properties Be Used in Conceptual Modelling? A Theory and Three Empirical Tests", *Information Systems Research* (12:4) 2001, pp. 384-405.

Bohrnstedt, G.W. "Reliability and Validity Assessment in Attitude Measurement", in: *Attitude Measurement,* Summers, G.F. (ed.), Rand-McNally, Chicago, Illinois, 1970, pp. 80-99.

Boudreau, M.-C., Gefen, D., and Straub, D.W. "Validation in Information Systems Research: A State-of-the-Art Assessment", *MIS Quarterly* (25:1) 2001, pp. 1-16.

Bowen, P.L., O'Farrell, R.A., and Rohde, F. "Analysis of Competing Data Structures: Does Ontological Clarity Produce Better End User Query Performance", *Journal of the Association for Information Systems* (7:8) 2006, pp. 514-544.

BPMI.org and OMG *Business Process Modeling Notation Specification. Final Adopted Specification*, Object Management Group, 2006 (available online at http://www.bpmn.org).

Bunge, M.A. *Treatise on Basic Philosophy Volume 3: Ontology I - The Furniture of the World*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1977.

Burt, R.S. "Interpretational Confounding of Unobserved Variables in Structural Equation Models", *Sociological Methods and Research* (5:1) 1976, pp. 3-52.

Burton-Jones, A. and Meso, P. "Conceptualizing Systems for Understanding: An Empirical Test of Decomposition Principles in Object-Oriented Analysis", *Information Systems Research* (17:1) 2006, pp. 38-60.

Cohen, J. "A Coefficient of Agreement for Nominal Scales", *Educational and Psychological Measurement* (20:1) 1960, pp. 37-46.

Davis, F.D. "A Technology Acceptance Model for Empirically Testing New End-User Information Systems: Theory and Results", *Doctoral Thesis*, Massachusetts Institute of Technology, Boston, Massachusetts, 1986. Available from

Davis, F.D. "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology", *MIS Quarterly* (13:3) 1989, pp. 319-340.

de Vaus, D.A. *Surveys in Social Research*, (5th ed.), Taylor & Francis Books, London, England, 2001.

Downs Jr., G.W. and Mohr, L.B. "Conceptual Issues in the Study of Innovation", *Administrative Science Quarterly* (21:4) 1976, pp. 700-714.

Fornell, C. and Larcker, D.F. "Evaluating Structural Equations with Unobservable Variables and Measurement Error", *Journal of Marketing Research* (18:1) 1981, pp. 39-50.

Froehle, C.M. and Roth, A.V. "New Measurement Scales for Evaluating Perceptions of the Technology-Mediated Customer Service Experience", *Journal of Operations Management* (22:1) 2004, pp. 1-21.

Gemino, A. and Wand, Y. "Complexity and Clarity in Conceptual Modeling: Comparison of Mandatory and Optional Properties", *Data & Knowledge Engineering* (55:3) 2005, pp. 301-326.

Green, P. "Use of Information Systems Analysis and Design (ISAD) Grammars in Combination in Upper CASE Tools - An Ontological Evaluation", in *Proceedings of the 2nd CAiSE/ IFIP8.1 International Workshop on Evaluation of Modeling Methods in Systems Analysis and Design*, University of Nebraska-Lincoln, Barcelona, Spain, 1997, pp. 1-12.

Green, P. and Rosemann, M. "Integrated Process Modeling. An Ontological Evaluation", *Information Systems* (25:2) 2000, pp. 73-87.

Green, P. and Rosemann, M. "Ontological Analysis of Integrated Process Models: Testing Hypotheses", *Australasian Journal of Information Systems* (9:1) 2001, pp. 30-38.

Green, P., Rosemann, M., and Indulska, M. "Ontological Evaluation of Enterprise Systems Interoperability Using ebXML", *IEEE Transactions on Knowledge and Data Engineering* (17:5) 2005, pp. 713-725.

Grover, V., Lee, C.C., and Durand, D. "Analyzing Methodological Rigor of MIS Survey Research from 1980-1989", *Information & Management* (24:6) 1993, pp. 305-317.

Guarino, N. and Guizzardi, G. "In the Defense of Ontological Foundations for Conceptual Modeling", *Scandinavian Journal of Information Systems* (18:1) 2006, pp. 115-126.

Jasperson, J., Carter, P.E., and Zmud, R.W. "A Comprehensive Conceptualization of Post-Adoptive Behaviors Associated with Information Technology Enabled Work Systems", *MIS Quarterly* (29:3) 2005, pp. 525-557.

Jöreskog, K.G. and Sörbom, D. *LISREL 8: User's Reference Guide*, Scientific Software International, Lincolnwood, Illinois, 2001.

Jöreskog, K.G., Sörbom, D., du Toit, S., and du Toit, M. *LISREL 8: New Statistical Features*, Scientific Software International, Lincolnwood, Illinois, 2001.

Landis, J.R. and Koch, G.G. "The Measurement of Observer Agreement for Categorical Data", *Biometrics* (33:2) 1977, pp. 159-174.

Lewis, B.R., Templeton, G.F., and Byrd, T.A. "A Methodology for Construct Development in MIS Research", *European Journal of Information Systems* (14:4) 2005, pp. 388-400.

Lyytinen, K. ""Ontological Foundations of Conceptual Modeling" by Boris Wyssysek - A critical response", *Scandinavian Journal of Information Systems* (18:1) 2006, pp. 81-84.

Maes, A. and Poels, G. "Evaluating Quality of Conceptual Modelling Scripts Based on User Perceptions", *Data & Knowledge Engineering* (63:3) 2007, pp. 769-792.

Milton, S. and Kazmierczak, E. "An Ontology of Data Modelling Languages: A Study Using a Common-Sense Realistic Ontology", *Journal of Database Management* (15:2) 2004, pp. 19-38.

Moody, D.L. "Theoretical and Practical Issues in Evaluating the Quality of Conceptual Models: Current State and Future Directions", *Data & Knowledge Engineering* (15:3) 2005, pp. 243-276.

Moody, D.L. and Shanks, G. "Improving the Quality of Data Models: Empirical Validation of a Quality Management Framework", *Information Systems* (28:6) 2003, pp. 619-650.

Moore, G.C. and Benbasat, I. "Development of an Instrument to Measure the Perceptions of Adopting an Information Technology Innovation", *Information Systems Research* (2:3) 1991, pp. 192-222.

Nunnally, J.C. and Bernstein, I.H. *Psychometric Theory*, (3rd ed.), McGraw-Hill, New York, New York, 1994.

Opdahl, A.L. and Henderson-Sellers, B. "Ontological Evaluation of the UML Using the Bunge-Wand-Weber Model", *Software and Systems Modeling* (1:1) 2002, pp. 43-67.

Parsons, J. and Cole, L. "An Experimental Examination of Property Precedence in Conceptual Modelling", in *Proceedings of the 1st Asian-Pacific Conference on Conceptual Modelling*, Australian Computer Society, Dunedin, New Zealand, 2004, pp. 101-110.

Parsons, J. and Cole, L. "What do the Pictures mean? Guidelines for Experimental Evaluation of Representation Fidelity in Diagrammatical Conceptual Modeling Techniques", *Data & Knowledge Engineering* (55:3) 2005, pp. 327–342.

Patel, A., Sim, M., and Weber, R. "Stakeholder Experiences with Conceptual Modeling: An Empirical Investigation", in *Proceedings of the International Conference on Information Systems*, Association for Information Systems, Helsinki, 1998, pp. 370-375.

Raduescu, C., Tan, H.M., Jayaganesh, M., Bandara, W., zur Muehlen, M., and Lippe, S. "A Framework of Issues in Large Process Modeling Projects", in *Proceedings of the 14th European Conference on Information Systems*, Association for Information Systems, Goeteborg, Sweden, 2006, pp. 1594-1605.

Recker, J. "Why Do We Keep Using A Process Modelling Technique?" in *Proceedings of the 18th Australasian Conference on Information Systems*, The University of Southern Queensland, Toowoomba, Australia, 2007, pp. 49-59.

Recker, J. "Understanding Process Modelling Grammar Continuance: A Study of the Consequences of Representational Capabilities", *Unpublished PhD Thesis*, Queensland University of Technology, Brisbane, Australia, 2008. Available from http://sky.fit.qut.edu.au/~recker/.

Recker, J., Indulska, M., Rosemann, M., and Green, P. "Do Process Modelling Techniques Get Better? A Comparative Ontological Analysis of BPMN", in *Proceedings of the 16th Australasian Conference on Information Systems*, Australasian Chapter of the Association for Information Systems, Sydney, Australia, 2005.

Recker, J., Indulska, M., Rosemann, M., and Green, P. "How Good is BPMN Really? Insights from Theory and Practice", in *Proceedings of the 14th European Conference on Information Systems*, Association for Information Systems, Goeteborg, Sweden, 2006, pp. 1582-1593.

Rockwell, S. and Bajaj, A. "COGEVAL: Applying Cognitive Theories to Evaluate Conceptual Models", in: *Advanced Topics in Database Research. Volume 4,* Siau, K. (ed.), Idea Group, Hershey, Pennsylvania, 2004, pp. 255-282.

Rohde, F. "An Ontological Evaluation of Jackson's System Development Model", *Australasian Journal of Information Systems* (2:2) 1995, pp. 77-87.

Rosemann, M. "Potential Pitfalls of Process Modeling: Part A", *Business Process Management Journal* (12:2) 2006, pp. 249-254.

Rosemann, M., Green, P., Indulska, M., and Recker, J. "Using Ontology for the Representational Analysis of Process Modeling Techniques", *International Journal of Business Process Integration and Management*) forthcoming, in press.

Rosemann, M., Vessey, I., and Weber, R. "Alignment in Enterprise Systems Implementations: The Role of Ontological Distance", in *Proceedings of the 25th International Conference on Information Systems*, Association for Information Systems, Washington D.C., 2004, pp. 439-448.

Segars, A.H. and Grover, V. "Re-Examining Perceived Ease Of Use And Usefulness: A Confirmatory Factor Analysis", *MIS Quarterly* (17:4) 1993, pp. 517-525.

Shanks, G., Tansley, E., Nuredini, J., Tobin, D., and Weber, R. "Representing Part–Whole Relations in Conceptual Modeling: An Empirical Evaluation", *MIS Quarterly* (32:3) 2008, pp. 553-573.

Sherif, M. and Sherif, C.W. "Attitudes as the Individual's own Categories: The Social-Judgment Approach to Attitude and Attitude Change", in: *Attitude, Ego-Involvement and Change,* Sherif, C.W. and Sherif, M. (eds.), Wiley, New York, New York, 1967, pp. 105-139.

Siau, K. and Rossi, M. "Evaluation Techniques for Systems Analysis and Design Modelling Methods - A Review And Comparative Analysis", *Information Systems Journal*) 2008, p. in press.

Sivo, S.A., Saunders, C., Chang, Q., and Jiang, J.J. "How Low Should You Go? Low Response Rates and the Validity of Inference in IS Questionnaire Research", *Journal of the Association for Information Systems* (7:6) 2006, pp. 351-414.

Stone, E.F. *Research Methods in Organizational Behavior*, Scott Foresman & Co, Santa Monica, California, 1981.

Straub, D.W. "Validating Instruments in MIS Research", *MIS Quarterly* (13:2) 1989, pp. 147-169.

Straub, D.W., Boudreau, M.-C., and Gefen, D. "Validation Guidelines for IS Positivist Research", *Communications of the Association for Information Systems* (13:24) 2004, pp. 380-427.

Sveiby, K.E. *The New Organisational Wealth: Managing and Measuring Knowledge-Based Assets*, Berrett-Koehler Publishers, San Francisco, California, 1997.

Wand, Y. "Ontology as a Foundation for Meta-modelling and Method Engineering", *Information and Software Technology* (38:4) 1996, pp. 281-287.

Wand, Y., Monarchi, D.E., Parsons, J., and Woo, C.C. "Theoretical Foundations for Conceptual Modelling in Information Systems Development", *Decision Support Systems* (15:4) 1995, pp. 285-304.

Wand, Y. and Weber, R. "An Ontological Model of an Information System", *IEEE Transactions on Software Engineering* (16:11) 1990, pp. 1282-1292.

Wand, Y. and Weber, R. "On the Ontological Expressiveness of Information Systems Analysis and Design Grammars", *Journal of Information Systems* (3:4) 1993, pp. 217-237.

Wand, Y. and Weber, R. "On the Deep Structure of Information Systems", *Information Systems Journal* (5:3) 1995, pp. 203-223.

Wand, Y. and Weber, R. "Research Commentary: Information Systems and Conceptual Modeling - A Research Agenda", *Information Systems Research* (13:4) 2002, pp. 363-376.

Weber, R. *Ontological Foundations of Information Systems*, Coopers & Lybrand and the Accounting Association of Australia and New Zealand, Melbourne, Australia, 1997.

Weber, R. "Conceptual Modelling and Ontology: Possibilities and Pitfalls", *Journal of Database Management* (14:3) 2003, pp. 1-20.

Weber, R. and Zhang, Y. "An Analytical Evaluation of NIAM's Grammar for Conceptual Schema Diagrams", *Information Systems Journal* (6:2) 1996, pp. 147-170.

Wyssusek, B. "On Ontological Foundations of Conceptual Modelling", *Scandinavian Journal of Information Systems* (18:1) 2006, pp. 63-80.